Application of the JDL Data Fusion Process Model for Cyber Security

Nicklaus A. Giacobe*
College of Information Sciences and Technology, The Pennsylvania State University
101 IST Building, University Park, PA 16802

ABSTRACT

A number of cyber security technologies have proposed the use of data fusion to enhance the defensive capabilities of the network and aid in the development of situational awareness for the security analyst. While there have been advances in fusion technologies and the application of fusion in intrusion detection systems (IDSs), in particular, additional progress can be made by gaining a better understanding of a variety of data fusion processes and applying them to the cyber security application domain. This research explores the underlying processes identified in the Joint Directors of Laboratories (JDL) data fusion process model and further describes them in a cyber security context.

Keywords: Cyber Security, Network Security, Data Fusion, JDL Data Fusion Process Model, Situational Awareness

1. INTRODUCTION

It was proposed[1] that data fusion techniques should be used to enhance situational awareness into network security events. However, since then, there have been few successes in adopting multi-sensor data fusion technologies for cyber security. This may be due to limited understanding of the entire data fusion process and it can and should be applied in cyber security. Researchers appear to be focused on using network-based intrusion detection systems and fusing their outputs to gain a more full understanding of undesired activities on the network. While there has been some success in this effort, overall awareness of the current status of the network and projection of future actions of adversaries has still not been achieved. The problem is much more complex and requires the fusion of data from widely varied sources, using multiple algorithms to achieve fusion and awareness at different levels and contexts. Most importantly, it must be understood that these tools are situational awareness aids. Their objective is to assist the human cyber security analyst and therefore must take into account human cognitive capabilities. The goals of this research, therefore, are to describe the JDL Data Fusion Process Model in cyber security terms (Section 2), categorize and describe the efforts of previous research in this domain (Section 3) and suggest areas for future work (Section 4).

2. THE JDL DATA FUSION PROCESS MODEL IN CYBER SECURITY TERMS

The JDL Data Fusion Process Model[2] is a reference model that describes the overall process of combining data from varied sources to result in better understanding of the situation being observed[3]. At each level of the fusion process, algorithms can be applied to combine the data and to make inferences about the meaning of the data in context. It is this inference capability that gives the data fusion system its power. At each level of the model, specific questions that can be answered in order to gain better situational awareness will be presented. The inference capability of specific implementations at each level will be described in cyber security terminology. It is the hope that this descriptive model will be used by future researchers to build new data fusion-driven cyber security tools. It is important to note that individual fusion applications do not have to address the JDL model in its entirety. Individual fusion applications can focus on identification, while others address future impacts, for example[2]. However, the entire scope of topics discussed in this paper need to be addressed by comprehensive systems of systems to facilitate situational awareness for security analysts.

Before describing cyber security functions at each level of the JDL Data Fusion Process Model, it is important to understand the relationship of the basic components and levels of the fusion process in cyber security terms (See Figure 1). Sensors are devices in the system that provide information about the system's security. Obvious examples are intrusion detection systems (IDSs), firewall logs and host-based security logs. The Level 0 fusion process would align data from various sources, addressing the problems of unique key matching (e.g. host name to IP address), chronometer

* nxg13@ist.psu.edu; phone 1 814 863 8555; http://*ist.psu.edu/nc2if/*

differences, differences in reporting output formats, etc. The Level 1 fusion process combines these data to identify individual security events, as they would be observed from multiple sensors. A Level 2 fusion process would ideally combine multiple individual entities to provide a current-state system perspective. A Level 3 fusion process would provide a capability to predict future states of the system (e.g. will the system be compromised by a specific attack?) or future actions of an attacker (e.g. what will the attacker's next step be?). Level 4 addresses the system's capability to task sensors and maintain their health. Updating IDS signatures would be an example of a basic Level 4 fusion process. Level 5 is the interface between the human analyst and the data fusion system.

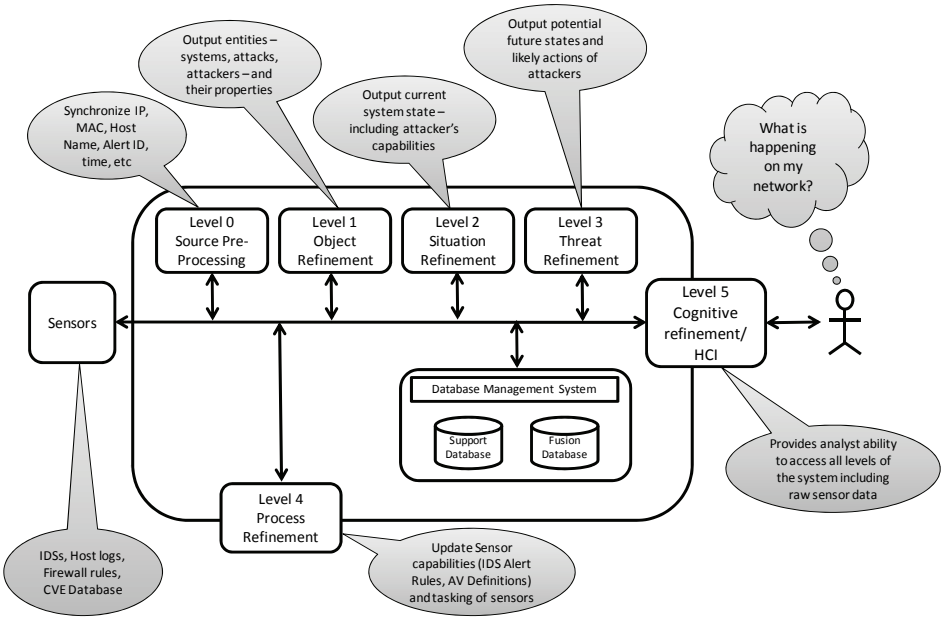

Figure 1. The Joint Director of Laboratories (JDL) data fusion process model adapted from Hall and McMullen (2004)[4]

2.1 Sensors

Sensors are devices that provide observation data for use in the data fusion system. In a cyber security system, it is obvious that network based IDS (NIDS) should be used, but they are not the only source of observations. Host-based IDSs are often overlooked or their output deprecated because of the likelihood of the software on the system being compromised and not being able to provide unbiased data. However, this does not mean that host-based sensors should not be used at all. In fact host log file data, system locations, operating systems installed, patches installed, system configuration and other routine IT-support data are all essential to understanding the current landscape of the network being defended.

Enterprise antivirus software applications can provide centralized logging and management of virus activities on hosts. Other enterprise management applications can be helpful as well in providing information about account login attempts, software (and versions) installed and other information that is critical for understanding the specific details on individual hosts on the network. Network devices are generally thought to be passive. However, they can provide specific information about which computing devices connected, disconnected and can even provide volume and characterization data of the traffic emanating from a given host. Switches and routers can provide data about IP addresses are in use by given hosts by matching them to MAC addresses. Firewalls could provide specific details about connection between hosts and the rest of the Internet, especially if configured to report on specific actions of interest taken by hosts.

Additional sensors might take the form of centralized knowledge bases. These could provide data about the availability of patches for the operating system, the list of known attacks and vulnerabilities, known tools and methods in use by attackers and other universally applicable data. Using these sources of data in different parts of the fusion system can provide a characterization of defended hosts, attacks, attack methods and possible future states.

2.2 Level 0/1 – Object Refinement/Data Alignment

These low level fusion processes identify, detect and characterize the individual entities[4]. In cyber security terms, these entities could be individual computers, adversaries, physical network connections, flows of data between hosts or other cyber entities. Individual computers are described with source data (from sensors, above) such as information about the system's physical location, operating system, hardware, patches, software installed, CPU utilization, security log, etc. Intrusions could be described with source data such as intrusion alert data from a NIDS like Snort, Cisco or others, host-based IDS information like security logs, anti-malware software, etc. Adversaries could be described as known or unknown, source locations, attack methods used, etc.

Table 1. Examples of Entities and Source Data

Entity	Source Data Examples
Defended Host	• Physical Location • Hardware Details • Operating System • List of Patches Applied • Application Level Software Installed • CPU and Memory Utilization • Applications Currently Running • End-User Account Access Data • Security Log Data • Classification of Data Stored on the System
Intrusion	• IDS Type and Location • Source Address • Destination Address • Attack Method • Time of Attack
Attacker	• Source Address(es) • Methods of Attack
Flows of Data	• Source and Destination Addresses • Type of Traffic (Protocols Used) • Traffic Volume • Encryption Used

The output of a Level 1 fusion system is a list of entities and their properties. The specific implementation of the fusion system could have a limited scope or focus. For instance, a fusion system might combine all of the known intrusion detection data from multiple IDSs and report a single intrusion (or attack) as being compromised of multiple IDS alerts. This correlation would aid the human analyst in being able to manage a limited number of attacks.

The types of algorithms that can be applied to combine sensor data in such a way as to provide entity identity are widely varied[4]. For cyber security, the issue is often resolved by matching hostnames, IP addresses and other similarly known data because different systems may report their findings based on different key factors. However, some cyber security systems provide data that is not 100% confirmed. For example, IDSs provide alert data that is plagued by high false positive rates and with lower successful detection rates. Because some IDSs have different capabilities – including pattern matching and anomaly detection methods –outputs from multiple IDSs could be fused to provide higher reliability and lower false positive rates. Rapid pattern matching/data mining techniques such as neural networks and support vector machines could be used to classify similar types of data from different classifying engines. These rapid, automated techniques can be helpful in fusing high volume, low reliability data from multiple different systems to provide a better, more reliable output. Figure 2 provides an example of such a system based on reports of research presented in Section 3.1 of data fusion using multiple IDSs.

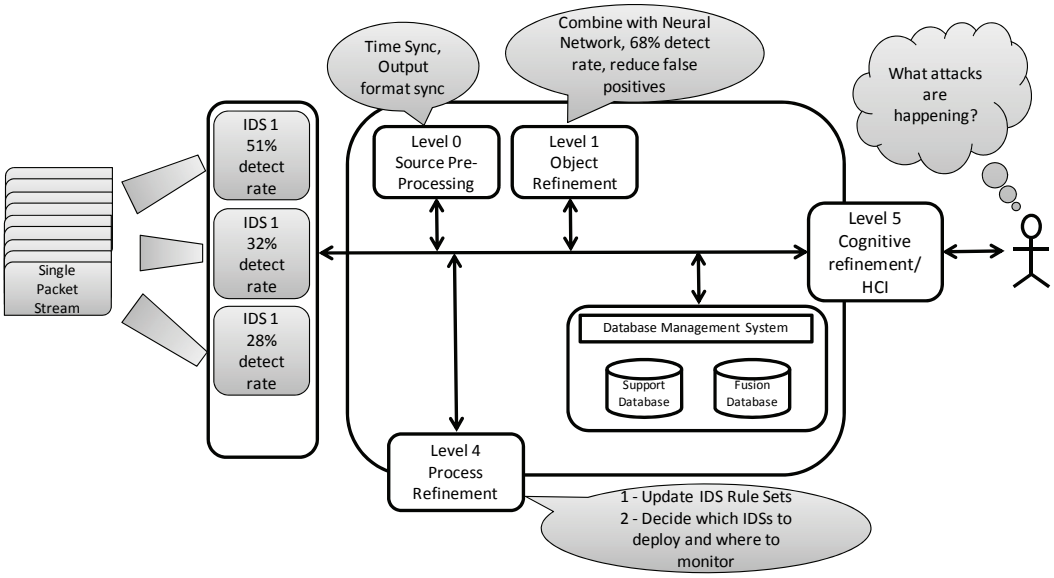

Figure 2. A Level 1/Object Refinement example of multiple IDSs monitoring same network traffic

2.3 Level 2 – Situation Refinement: Awareness

A Level 2 fusion process aggregates individual entities from lower level fusion processes in an attempt to describe the current system state. Entities are complex and have many features, but it is only these individual features that are detected by a specific sensor. The Level 1 process could potentially combine the multiple features of an individual entity, but the Level 2 process needs to combine those entities into a comprehensive picture of the current situation.

Combining data from multiple systems could provide a better understanding of the current state of the system. For example, combining a number of system health data such as operating system patches installed, antivirus software definition set being used, list of processes running on the system, and other similar data could provide an estimation of the system's overall health and ability to defend itself if attacked with a certain attack method. Representing all of the data for all systems (computers, servers, firewalls, VPN appliances, etc.) in defended network could provide the analyst with an awareness of the current system state. Deviations from that known state may prompt the analyst to investigate.

The algorithms that should be able to be applied to support Level 2 fusion are related to pattern matching and automated reasoning[4]. The reports from a system on its own health (patch level, antivirus definitions, etc) can be matched against a known, desired state. Systems that do not match the desired state can be flagged as being out of date, running undesired software or versions, or exhibiting undesired behaviors (high CPU load, low available drive space, high overall network traffic, etc). A host on the network could be further described as being in a specific location, having a network address, running a specific version of an operating system, having specific patches installed, running a version of an antivirus program with a known set of AV definitions, having specific application software installed, etc. Combining similar output results from all of the network's hosts, similar data from network devices such as routers and firewalls could be combined into a total view of the network's defensive posture. However, this is not sufficient to describe an entire network security state. There must be an understanding of the attacker's potential capabilities as well. Combining the attacker's capability set with the known state of the defensive posture could result in an awareness of the current system security level.

Therefore, to have awareness of a "situation", the output of a Level 2 fusion process must provide information about the defensive posture of the network under attack, the attacker's capabilities, the attacks that have occurred and an assessment of whether those attacks have been successful or not. Inferences about whether the attacks were successful (or not) could be developed based on further evaluation of the host that was attacked. For example: Are there new processes running since the attack? Did the host's security log provide additional relevant details? Is the host's configuration different than before? Did the host initiate communications to known or suspected bad sites? More proactive measures could compare the host's defensive capabilities vs. the attack's methods. For instance, if the host does not have the patch that addresses the vulnerability that the attacker uses, it could be assumed that the host was

compromised. Additionally, other sensors could provide additional information on exfiltration activities. For instance, Data Leak Protection (DLP) sensors could be used to identify whether the intruder has attempted to access or send sensitive data from the protected host. These inferences inform the analyst as to what to do next. If a host has been compromised, the analyst may decide to remove the host from the network and perform some kind of forensic analysis. Additionally, if a given host has not received or applied updates, this could be an indicator of a management system failure which could indicate policy or workflow related actions that need attention.

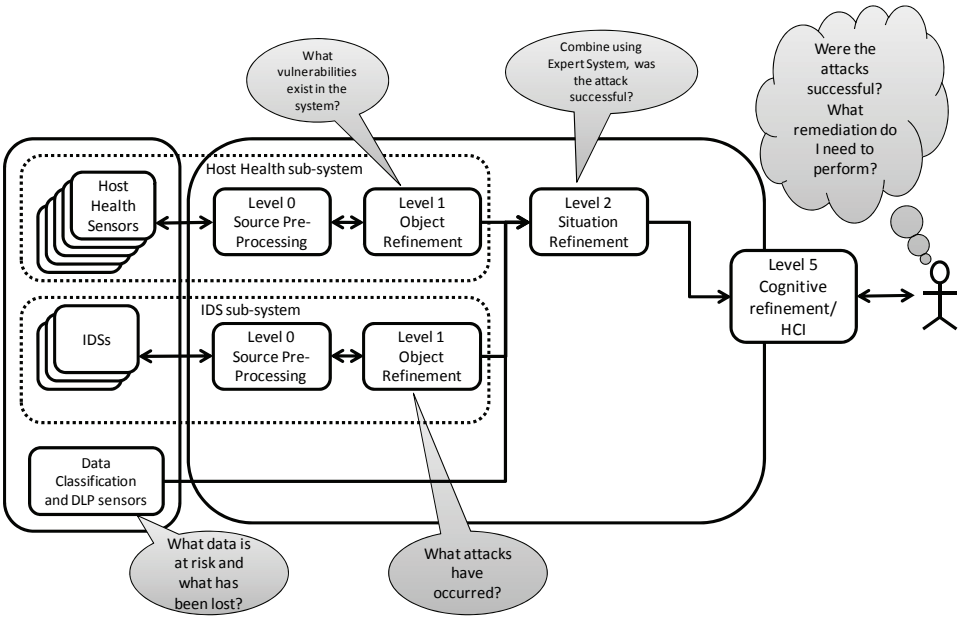

Figure 3. A Cyber Security Level 2/Awareness Fusion System Example

2.4 Level 3 – Threat Refinement: Projection

Knowing that a single host on the network has been compromised by a given attack method could inform the analyst of remediation actions that need to take place elsewhere on the network – such as needing to apply firewall rules in order to work around a zero-day exploit until patches can be rolled out. Simply having knowledge of an attacker's possible capabilities (knowing that a specific zero-day exploit is in the wild) could trigger a similar response. These responses are projection of possible future states by the analyst who uses the output from the Level 2 fusion system and projects possible future states on his/her own.

Using a repository like the CVE Database[5] may help to build an understanding of the attacker's options. However, data from these systems need to be coordinated with the defensive posture of the defended network. Simply knowing that an attacker has an attack option is not sufficient without knowing what is likely to happen if that attack is used on the given defended network with a known set of defensive capabilities (patches, firewalls, IDS/IPS actions, etc). Additionally the data from the CVE database is so vast and changes so quickly that attempting to process it manually by a human analyst would be overwhelming. Some kind of automated fusion process can combine an understanding of the capabilities of the attack referenced in the CVE database and compare it to the current set of hosts in the network and their "health". From this combination, a list of exploits for which the hosts are vulnerable could be created. From that list of vulnerabilities, additional defensive system capabilities could be evaluated to determine if an attacker could be successful with that particular attack method. The analyst could decide if there are additional actions that need to be taken – patches to apply, firewall rules to configure, IDS/IPS actions to take, etc. A decision of which actions to take first could be informed by an understanding of the data that is at risk (from data classification sensors) on given hosts.

It is also helpful to understand the capabilities of the attacker from a tool perspective. The analyst should be aware of the tools that are at the disposal of the attacker and how those tools are updated. For example, knowledge of the tools and methods of attackers, and how those tools integrate new exploits as they become available would be helpful to understanding the capabilities of an attacker. Other attack tools could also be monitored for new updates and a fusion system could be developed to provide a Level 0/1 assessment of attack options available to the attacker. This kind of

assessment could provide an understanding of what the attacker will do next. While a novice attacker might simply use every attack option available, a more sophisticated attacker might try rudimentary attacks and then choose new attack methods based on the feedback received from initial attempts.

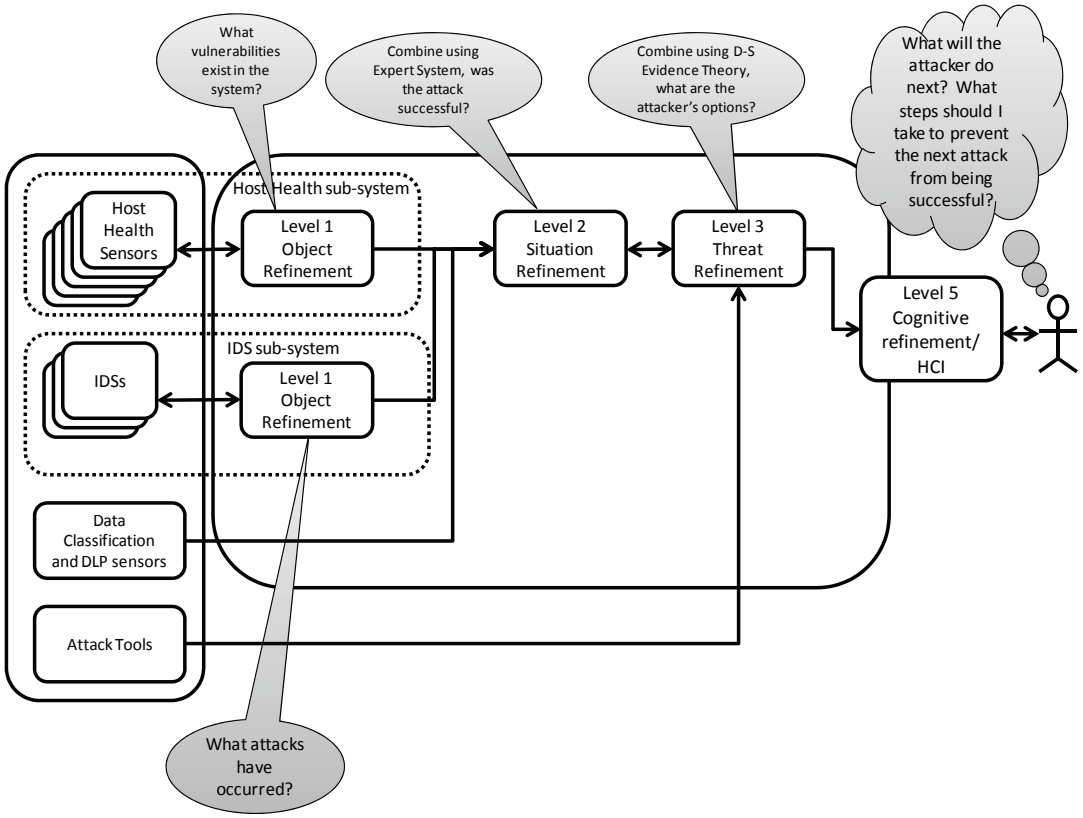

Figure 4. A Cyber Security Level 3/Projection Fusion System Example

2.5 Level 4 – Process Refinement: Sensor Management and Selection

Hall describes L4 as a metaprocess, tasked with observing the fusion system and taking input from outside of that system[4]. In terms of classic examples of military fusion processes, deciding where to "look" and which sensors to use may be dictated by the mission objectives. The use of certain technologies in a military application (radar, sonar, etc.) may be contraindicated in certain missions or parts of the mission (where stealth is a mission requirement, for example)[4]. In the cyber domain, detection technologies are not as controllable as physical sensors are. However, providing sensors with the capabilities to detect new attack methods (through IDS rule set and antivirus definition updates for example) could be considered as example of a sensor tasking role.

The remainder of Level 4 is a decision process on where to look and what to look for. There are constraints in defensive cyber security regarding the use of passive tools vs. active ones. However, these constraints are legal and political, not technical. Under proper legal jurisdictional authority, it may be possible to employ back hacking to provide additional information about the attacker, by taking control of one of the attacker's bots and observing their command and control functions.

Even in a purely defensive operational mode, the selection of which tools to use and what to monitor for represents an L4 function with a human-in-the-loop providing mission requirements. An example of this would be monitoring for specific types of traffic (the presence of IRC traffic, or connections to servers/hosts in foreign countries for example) as an indicator of compromised hosts on the network. If a host on the network starts opening a number of connections to computers on the Internet on Port 25, a border monitoring service could alert an analyst of a possible spam-bot infection on that host. As new compromise methods are found, specific monitoring tools can be deployed by the analyst to observe evidence of those compromises. In cyber security today, these L4 processes are generally manual and therefore slow to deploy.

2.6 Level 5 – Cognitive Refinement and HCI

As compared to data fusion in other domains, network security analysts are burdened with two confounding issues. The first is that high rate of data increasing cognitive load. Many current cyber security systems provide data to the analyst in text-based formats. As false alert rates in some cyber security systems are extremely high, these overwhelm the analyst and result in real intrusion and attacks going unnoticed. Cognitive Load Theory identifies the number of items that an individual can handle in working memory. Working memory is used to access long-term memory, which is thought to be organized in schemas or structures[6]. Seasoned analysts are capable of a higher level of understanding of what is normal and abnormal on networks that they monitor by matching current scenarios to previously-built schemas[7]. However, schema building is time consuming.

In the 1999 revision of the JDL Data Fusion Process Model diagrams, the HCI part of the system appears to accept inputs from the highest level of the fusion process[3]. However, the description of the Level 5 fusion process has been clarified to indicate that the HCI interface should provide access to and human control at each level of the fusion process[8]. The design mantra of "Overview, Filter/Zoom and Details-on-Demand" provides a suggestion for how to organize and display data in an information visualization system[9]. Following this mantra, HCI for a network security data fusion system should present the security analyst with an overview of the defended system, likely of current system state. The analyst should have the ability to zoom into specific region of the defended network (even to the level of an individual host), select specific details from the raw data, and filter them to select specific details. This spans all of the lower levels (0-3) of the data fusion system. The analyst should have the ability to select, filter and zoom in on specific details. This level of detail capability is an important sensemaking feature for the analyst's cognitive schema creation and maintenance processes as well as to support forensic and legal reporting requirements.

This brings to light the second issue in Level 5 in cyber security, the lack of a shared mental model of the problem space. Because the "terrain" is virtual, an analyst may have a different mental picture of what the defended network looks like than the data fusion system designer, or even other security analysts. While a representation of the defended network may take the form of the physical terrain, there really is no physical space constraint. Often, logical topologies are more suitable to representing the workspace. It should be noted that a single, optimal representation is unlikely to be created. The goal of L5 should be to present the data in such a way as to leverage the analyst's understanding of the cyber workspace and allow the analyst to access security information at all levels of the system.

3. ORGANIZING PREVIOUS RESEARCH

3.1 Level 0/1 IDSs and Real-Time Fusion Algorithms

One of the primary concerns about IDSs is that they have a high rate of false positives. Additionally, alerts from an IDS can often represent part of an attack, while other alerts may represent other parts of the same attack. A correlation algorithm could merge multiple alerts that have a common source, destination or attack method. However, the use of automated attack tools (like the Metasploit framework, for example) allow unsophisticated attackers the capability to launch many attempts using different methods on multiple destinations. Bot networks are also employed by slightly more organized attackers to mask the true source of the attack by launching from compromised hosts instead of from the attacker's true location. These methods result in IDS alerts that are difficult to correlate by simply matching source, destination or method.

A number of researchers have tested proposed solutions by evaluating them against the DARPA Intrusion Detection Data Sets from 1998-2000[10]. These standard data sets provide a known, but limited, set of attacks in progress in a large corpus of normal network traffic. Liang et al. proposed the use of neural networks to analyze network data from the 1998 DARPA data set in order to extract features from the data and compare the fitness of their neural network-based approach to genetic algorithms. An overall security risk value is calculated, providing an overall assessment of the network security situation on a scale of [0,1] [11]. However, other than the assessed value, no information about the underlying reasons for the increase or decrease in the assessment is provided.

Wang et al. provided more description in their use of neural networks. In the combination of data from a Snort IDS and from a Cisco Netflow collector, they identified that some feature reduction was required as to not overload the neural network with unnecessary data fields from each of the IDSs. They proposed the use of a multi-layer feed-forward neural network (MLF-NN). Using the DARPA 1999 data set, they were able to train and evaluate their algorithm. Their output

provided a rapid fusion and assessment of the intrusion behavior identified in the traffic. However, their results showed a higher than desired false positive rate, implying that more sensors are required to eliminate the false alerts [12]. The same research team also developed a prototype of a multiclass support vector machine-based (SVM) fusion engine and reported their findings in[13, 14]. Their assessment was that the SVM approach was just as reliable, but much faster, indicating that SVMs have promise in real-time IDS applications.

Thomas and Balakrishnan developed a data fusion system based on neural networks that combined the results of three different IDS systems that monitored the DARPA 1999 data set. The individual capabilities of PHAD, ALAD and Snort IDS systems had a success rate of 28%, 32% and 51% respectively. However by fusing the alert data from all of the systems together, they were able to increase to the total successful detection rate to 68% and significantly increase the detection rate of certain attack types even though those attacks were not detected well by each individual IDS [15, 16]. It is this work that is represented in Figure 2, above.

Conceptually, fusing the outputs from different IDS systems with different detection capabilities in order to increase detection rates and decrease false positive notification rates is an exemplar of a Level 0/1 fusion algorithm's purpose. The output of this fusion process is still individual objects (the attacks), which are important elements for the next level of the fusion process as these and other elements from the environment are fused to provide an overall assessment of the current state of the system.

3.2 Level 2/Awareness and Level 3/Projection Systems

Yang et al. developed a system with different algorithms for Levels 2 and 3 of the fusion process. The first system, INFERD, provided automated alert correlation using *a priori* knowledge from subject matter experts on both the attack methods and knowledge of the specific vulnerabilities of the defended system. Fusion of these two data sets provided an assessment of the likely outcomes on the given attack on the given network, reducing the time for assessment by limiting the problem set to the specific problem space[17]. Yang chose to use *a priori* information about the attack methods and known vulnerabilities. A pattern-matching, data mining or other high speed comparison algorithm could have been used instead. However, this would have required significant development of the set of system states, attack methods and vulnerability combinations. The use of expert systems in this domain appears appropriate, but results are limited to the specific knowledge of the experts. Closing the loop on INFERD when used in an operational setting might include the use of a Data Leak Protection (DLP) monitoring system to identify when hosts have been compromised. This additional data input to the Level 2 fusion algorithm could provide a learning mechanism to confirm that compromises did occur.

The second system was the corresponding threat projection module, TANDI, which uses the attack tracks provided by INFERD and overlays an a priori model of the defended network to identify specific nodes (computers, accounts, databases, etc.) that are subject to compromise. Using knowledge of previous successful and unsuccessful attacks, future attacks can be probabilistically predicted. TANDI combines assessments of possible outcomes and threat levels based on a variety of factors including the intent and skill level of the hacker, value of the potential target and the utility of the vulnerability being used. These provide potentially conflicting predictions and TANDI combines them using a Dempster-Shaffer Theory-based combination[18]. D-S Theory is a Bayesian-like algorithm that allows for the combination of evidence without necessarily knowing all possibilities. A compromised host would receive a value of 1, while evidence of potential compromise can be given weighted scores [0,1] and combined using the Dempster Rule. TANDI was tested with some un-optimized weights in its algorithm, yet still provided reasonable assessments[17].

INFERD and TANDI were evaluated with extremely interesting and promising results. However, with these promising results comes a caveat – the threat assessment and prediction algorithms that are based on *a priori* information are susceptible to the limitations of the knowledge of the experts who provided that information[17]. Flaws in the basic underlying knowledge will result in flaws in the output. In particular, it appears that while these are excellent systems for the detection of known methods of attack on a known infrastructure, the analyst-in-the-loop having an understanding of the defended network and human instinct to identify the abnormal behavior in a complex system a critical component. Systems like INFERD and TANDI are certain to be extremely beneficial in the development of awareness for the human analyst.

Virtual Terrain (VT) addresses the question of whether an attempt to compromise a given host by only representing the relevant information and presenting that to an analyst[19]. The automated process maps hosts, processes, services, users, vulnerability possibilities and other important details on to a logical map of the hosts on the network. Using the VT as a

basis, attacks are evaluated to determine their success based on the network (routers, switches, firewalls, etc) and host configurations (host OS vs. vulnerability match). Other methods that could be used are attack graphs and vulnerability trees. However, these modeling mechanisms require significant human expert involvement to build them and are computationally intensive to evaluate. Therefore, these tools may be difficult to implement in enterprise architectures or for large numbers of hosts.

3.3 HCI and Level 5 Solutions

Cyber security systems have been plagued by limited HCI development. Commonly hosts and network devices are represented on maps – either logical or physical and connected via line segments to indicate physical network connections. However, these types of representations do not scale to large networks. Others have created representations of network traffic between hosts and networks using network diagrams and treemaps[20]. However, most production systems rely on textual output – from IDS alerts, firewall logs, host security logs and other similar data is simply represented as lines and lines of text.

4. FUTURE WORK

The work of implementing data fusion capabilities in the cyber security domain is still in its infancy. This paper is intended to be a description of the theoretical model in terms that cyber security researchers will find more familiar. This explanation of the model will need additional revision to include other areas of cyber security that are not addressed in this paper. This work is predicated on the organizational level of the problem space, but individuals and small networks have their own needs that are not addressed in this paper. However, even in this organizational level assessment, more work needs to be done in several areas of development including sensors, algorithms and HCI.

4.1 Sensor Development

Additional sensors need to be developed to provide evidence of cyber security status. Sensors could be simple, ad-hoc software tools on each system in a defended network. They could take on the form of complex intrusion detection and prevention systems. However, a number of systems already exist in the enterprise environment that could be leveraged as sensors. The most obvious include antivirus systems, centralized management systems (Microsoft SMS, for example), patch management systems, centralized authentication systems (Active Directory, OpenLDAP, etc), host log files and other similar tools that exist in the enterprise, but remain untapped for fused cyber security systems.

4.2 Algorithm Development

Algorithms need to be developed at all levels of the JDL model. The hints as to which algorithms might be most fruitful in specific situations should be taken from other data fusion domains. New algorithms don't need to be invented as much as algorithms need to be selected and applied to the problem space of cyber security. The high data rate of elements in cyber security might pose unique problems and limitations, but algorithm choice for lower vs. higher levels of fusion could be guided by successes in other domains.

4.3 Level 5/HCI

The human-computer interaction aspect of the fusion process needs the most attention. The analyst needs access to all levels of the fusion process, not simply the highest level available. Awareness of the individual entities in the situation and their properties requires that the analyst have the ability to access the system from the highest visual representation down to the lowest data element available. These qualities are also important for the analyst's ability to respond to legal requirements and requests forensic analyses and details. However, the analyst has only a limited (human) capacity to observe and interpret these details. New HCI development needs to take the analyst's cognitive capabilities into consideration to reduce fatigue and support attention to the important details of security situation.

5. CONCLUSION

Multi sensor data fusion is a powerful capability that can provide significant advances in cyber security, especially from an organization-level perspective. Advances in the application of fusion concepts in cyber security can be represented in all levels of the JDL Data Fusion Process Model. New contributions in cyber security data fusion can provide for entity identification, support of situational awareness and projection of possible future states or adversary actions. This work may take the form of sensor development, algorithm application or system-to-human interface design. This work must take into account that the awareness is a human cognitive process that needs technological support from fusion systems. New contributions in HCI/Level 5 fusion processes is an important next step in fusion-based cyber security systems.

REFERENCES

[1] Bass, T., "Intrusion detection systems and multisensor data fusion," Commun. ACM, 43(4), 99-105 (2000).
[2] Kessler, O., Askin, K., Beck, N., Lynch, J., White, F., Buede, D., Hall, D., and Llinas, J., [Functional description of the data fusion process] Office of Naval Technology, Naval Air Development Center, Warminster, PA(1991).
[3] Steinberg, A., White, F., and Bowman, C., "Revisions to the JDL data fusion model." Proc. SPIE, Vol. 3719, 430-441 (1999).
[4] Hall, D., and McMullen, S., [Mathematical techniques in multisensor data fusion] Artech House, Boston, MA(2004).
[5] [CVE - Common Vulnerabilities and Exposures (CVE) http://cve.mitre.org/].
[6] Sweller, J., Van Merrienboer, J., and Paas, F., "Cognitive architecture and instructional design," Educational psychology review, 10(3), 251-296 (1998).
[7] D'Amico, A., and Whitley, K., "The Real Work of Computer Network Defense Analysts," Proceedings of the Workshop on Visualization for Computer Security, VizSEC 2007. 19-37 (2007).
[8] Blasch, E. P., and Plano, S., "JDL level 5 fusion model: user refinement issues and applications in group tracking." Proc. of SPIE, Vol. 4729, 270-279 (2002).
[9] Shneiderman, B., "The eyes have it: A task by data type taxonomy for information visualizations," 1996 IEEE Symposium on Visual Languages. 336-343 (1996).
[10] [DARPA Intrusion Detection Data Sets] MIT Lincoln Laboratory: Information Systems Technology, Cambridge, MA(2008).
[11] Liang, Y., Wang, H.-Q., and Lai, J.-B., "Quantification of Network Security Situational Awareness Based on Evolutionary Neural Network," 2007 International Conference on Machine Learning and Cybernetics. 6, 3267-3272 (2007).
[12] Wang, H., Liu, X., Lai, J., and Liang, Y., "Network security situation awareness based on heterogeneous multi-sensor data fusion and neural network," Second International Multi-Symposiums on Computer and Computational Sciences, 2007. IMSCCS 2007. 352-359 (2007).
[13] Liu, X., Wang, H., Lai, J., Liang, Y., and Yang, C., "Multiclass Support Vector Machines Theory and Its Data Fusion Application in Network Security Situation Awareness," International Conference on Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. 6349-6352 (2007).
[14] Liu, X.-W., Wang, H.-Q., Liang, Y., and Lai, J.-B., "Heterogeneous Multi-Sensor Data Fusion with Multi-Class Support Vector Machines: Creating Network Security Situation Awareness," 2007 International Conference on Machine Learning and Cybernetics. 5, 2689-2694 (2007).
[15] Thomas, C., and Balakrishnan, N., "Advanced sensor fusion technique for enhanced Intrusion Detection," IEEE International Conference on Intelligence and Security Informatics. 173-178 (2008).
[16] Thomas, C., and Balakrishnan, N., "Performance enhancement of Intrusion Detection Systems using advances in sensor fusion," 11th International Conference on Information Fusion. 1-7 (2008).
[17] Yang, S., Stotz, A., Holsopple, J., Sudit, M., and Kuhl, M., "High level information fusion for tracking and projection of multistage cyber attacks," Information Fusion, 10(1), 107-121 (2009).
[18] Shafer, G., [The Dempster-Shafer theory], (1992).
[19] Holsopple, J., Yang, S., and Argauer, B., "Virtual terrain: a security-based representation of a computer network," Proc. of SPIE, Vol. 6973, 69730E (2008).
[20] Mansmann, F., and Vinnik, S., "Interactive Exploration of Data Traffic with Hierarchical Network Maps," Visualization and Computer Graphics, IEEE Transactions on, 12(6), 1440-1449 (2006).